STATISTICAL EDGE DETECTION

Scott Konishi, Alan Yuille, James Coughlan and Song Chun Zhu

Statistical Image Regularities

- 1. There are considerable statistical regularities in real images. (Field, Atick, Bialek, Ruderman, Simoncelli, Zhu, Mumford, … Green.)

- 2 Histograms of differential filters are very similar between images.

Edge Detection

- There have been a thousand PhD theses on edge detection (computer vision myth).

- None work significantly better than Canny’s master thesis (MIT 1983).

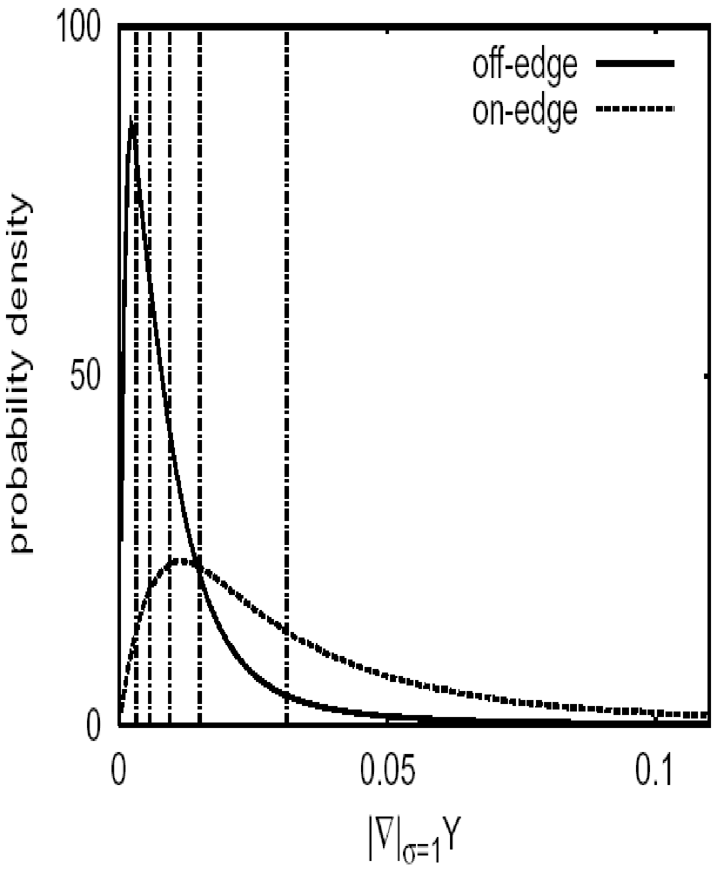
Not considering global methods such as Geman & Geman, Mumford, Osher, Zhu…

Statistical Edge Detection

- 1. Let $f(I(x))$ denote the filter response at point x on image I.
- 2. Let $P(f=y|x\ ON)$ and $P(f=y|x\ OFF)$ be the empirical distributions of the filter response, conditioned on x being ON or OFF an edge
- 3. Use loglikelihood ratio test to detect edges: $\log P(f=y|x\ ON)/P(f=y|x\ OFF) > T$.

Example

- Let f(I(x)) = |grad I(x)|
- Calculate empirical histograms P(f=y|ON) and P(f=y|OFF).
- P(f=y|ON)/P(f=yOFF) is monotonic in y.
- So loglikelihood test is threshold on |grad (I(x)|.



Coupling scalar filters

- Couple different edge cues by making f(.) vector-valued.

- Example, combine filters at different scales -- |grad G_sig * I|, where G_sig is a Gaussian with s.d. sig and * is convolution.

- Example, combine different filters at different colour bands.

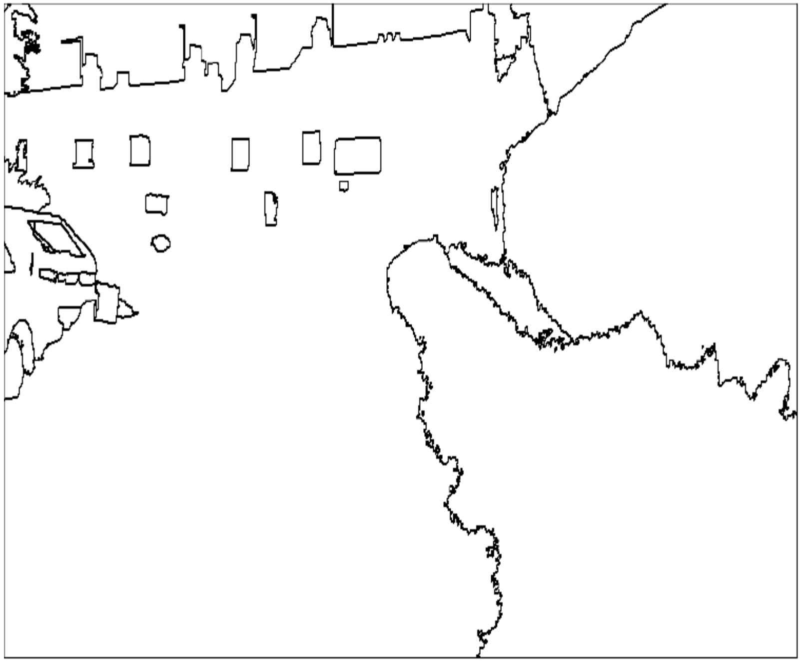
Datasets with Ground-Truth

- Sowerby Dataset – 100 colour images of English country with segmentations.

South Florida Dataset – 50 grey-scale Images with segementations.

Berkeley Dataset – 100’s of segmented images. People’s judgements of edges are very similar.

Sowerby Example



Representations

- Use non-parametric representations of the histograms/probability distributions.

Problem – the number of bins increases exponentially with the dimension of the filter.

The amount of training data must grow exponentially to ensure generalization.

Example

- A 9-dim filter with 10 bins per dimension has 1,000,000,000 bins.

- But 100 images with 500 x 800 pixels (each) has approximately 2,800,000 edges (7% per image).

Not enough data.

Our Strategy

- Adapt the representation to the amount of data available. Use cross-validation to check for overlearning.
- Select histogram bin boundaries for 1-dim filter to maximize performance measures (6 bins is adequate)

Use same bins for multi-dimensional filters AND use decision cuts (if necessary) to Reduce the representation.

Performance Measures

- ROC curve – plot false +'ves against false -ve's of loglikelihood test as threshold varies.
- Area under the ROC curve (error of two-alternative forced choice). Bayes risk,

Chernoff information – Bhattarcharyya bound. Motivated by theoretical studies by Yuille and Coughlan (2000).

All measures gave equivalent results.

Chernoff and Bhattarcharyya

The Chernoff Information between distributions $p(y)$ and $q(y)$ is :

$$C(p,q) = -\min_{0 \le \lambda \le 1} \log\{\sum_{y=1}^{J} p^{\lambda}(y) q^{1-\lambda}(y)\}.$$

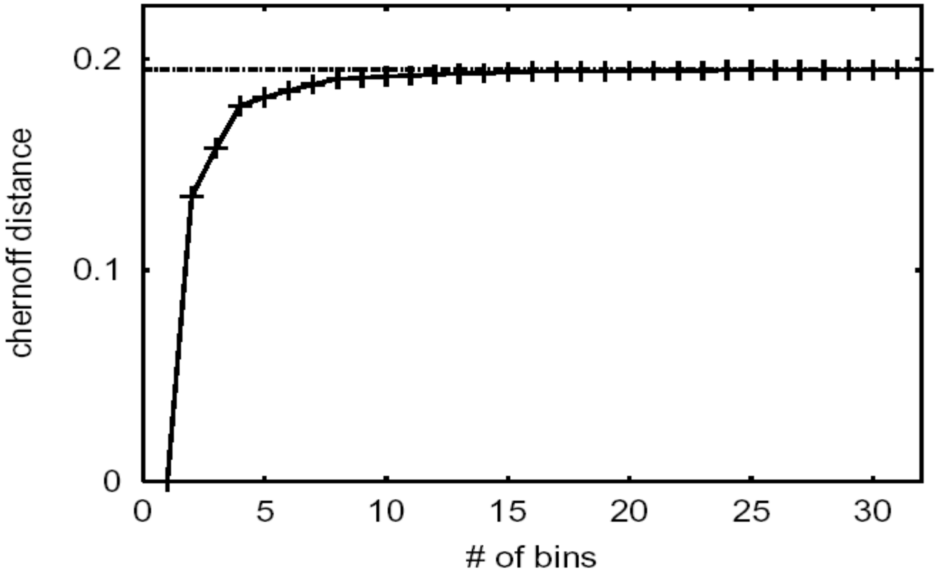
Motivated by order parameter theory for curve detection (Yuille, Coughlan 2000).

The Bhattarcharyya coefficient is :

$$B(p,q) = -\log\{\sum_{y=1}^{J} p^{1/2}(y) q^{1/2}(y)\}$$

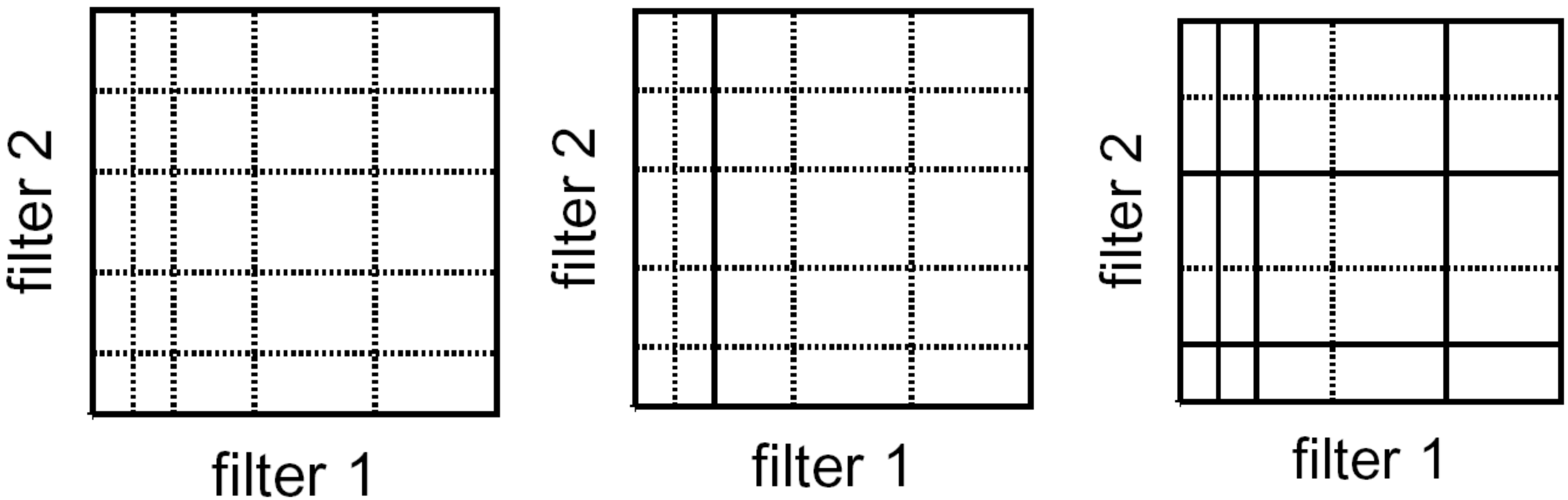
Choice of 1-D Bins

- Select bin boundaries to maximize Chernoff as a function of no. bins.
- |grad(I)|: C=0.125 for discrim. thresh.



Decision Tree Representation

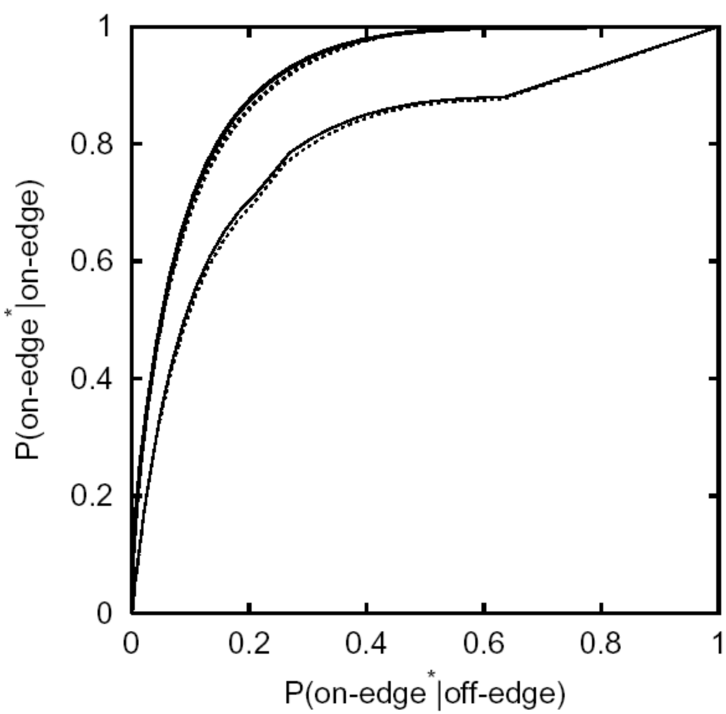
- Adaptively selects cuts on 1-D filter axes to maximize Chernoff. Compact representation requiring less data.

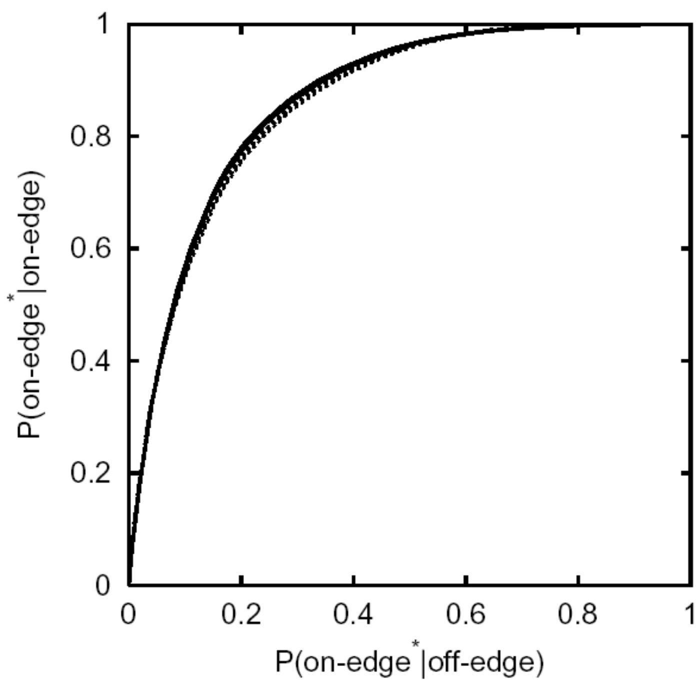


Cross-Validation

- Train on half dataset and test on rest.

Overlearning (left). True learning (right).



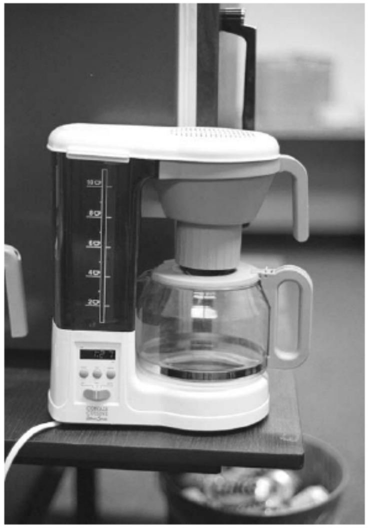


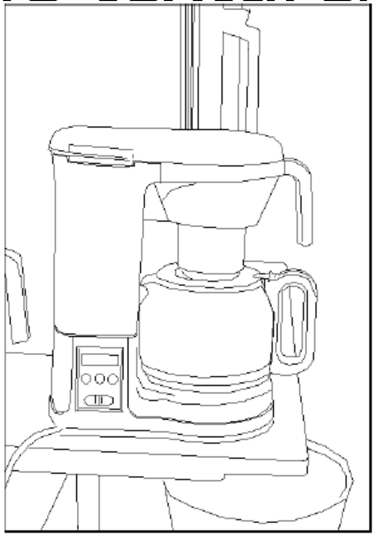
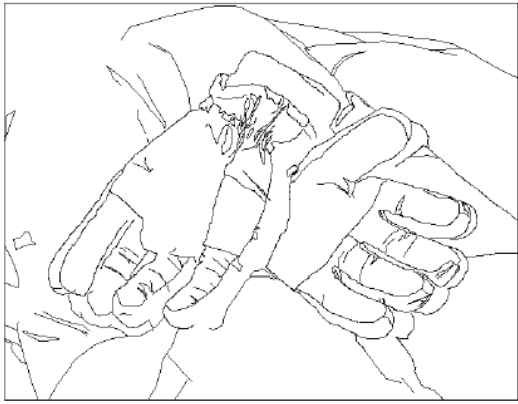
Two Datasets:I Sowerby

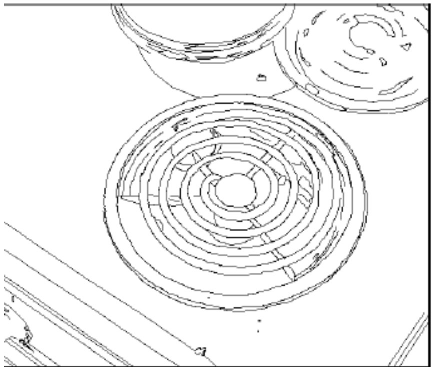
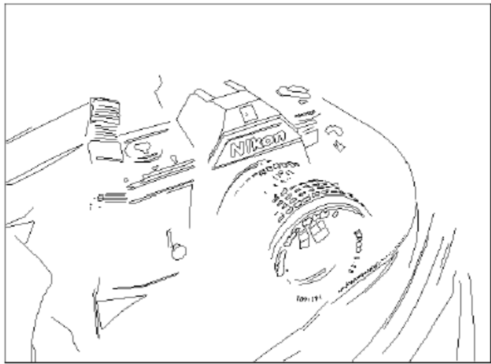
- Sowerby – much texture/clutter:

Two Datasets: II Florida

- South Florida: little texture/clutter



Filters

- Differential Operator: grad, Laplacian, Nitzberg, Gabors, Hilbert transform pairs.

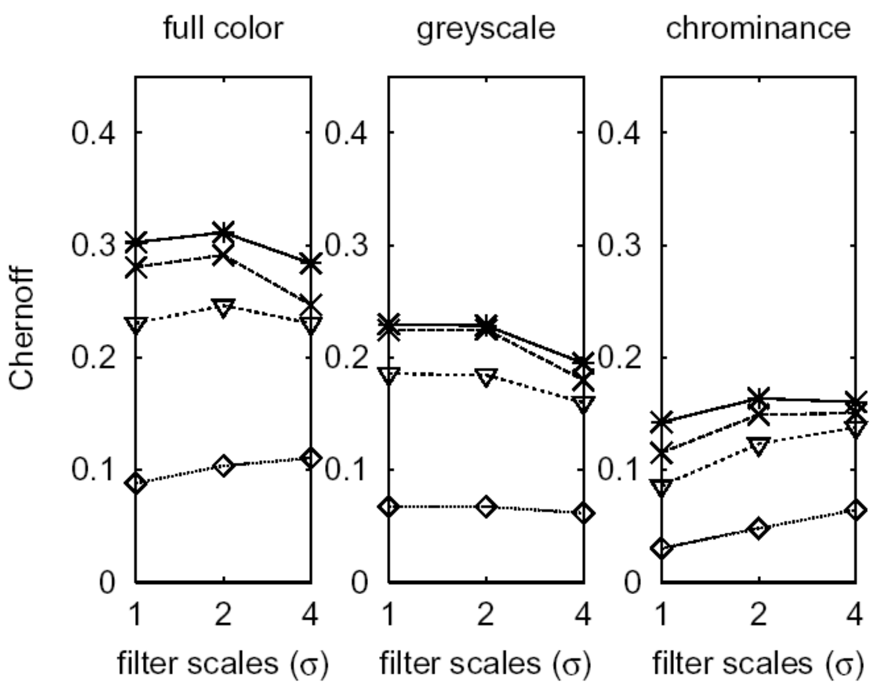
Scales: G(sig)*I: G Gaussian, sig SD, * Convolution.

Colour: Full colour, greyscale, chrominance

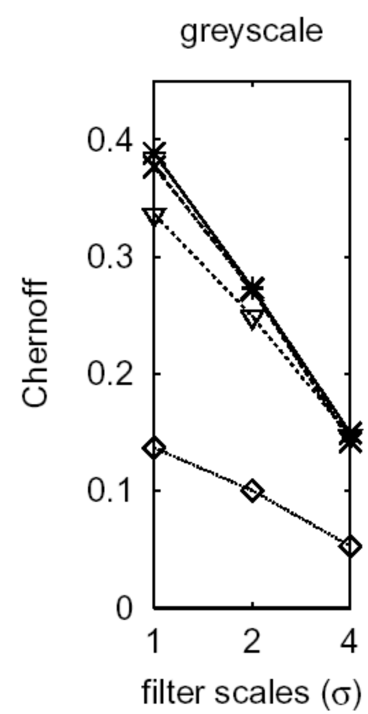
Filter Scales

- Sowerby (left), Florida (far right)

$$\text{Triangles}\,|\nabla|,\ \text{Diamonds}\,\nabla^2,\ \text{Stars}\ N_{1,2},\ \text{Crosses}\ N_1$$

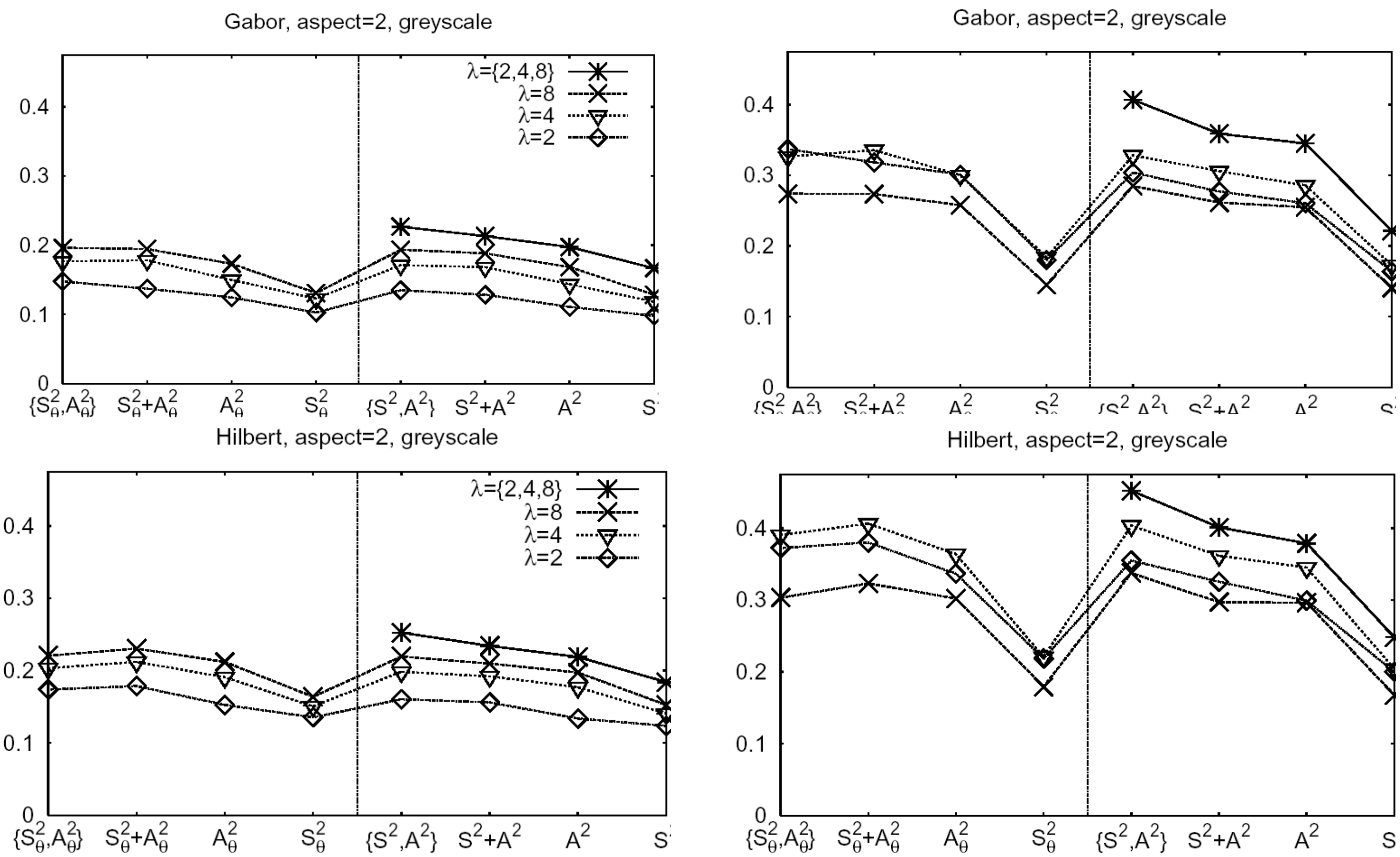


osses N_1



Oriented Filters:Biology?

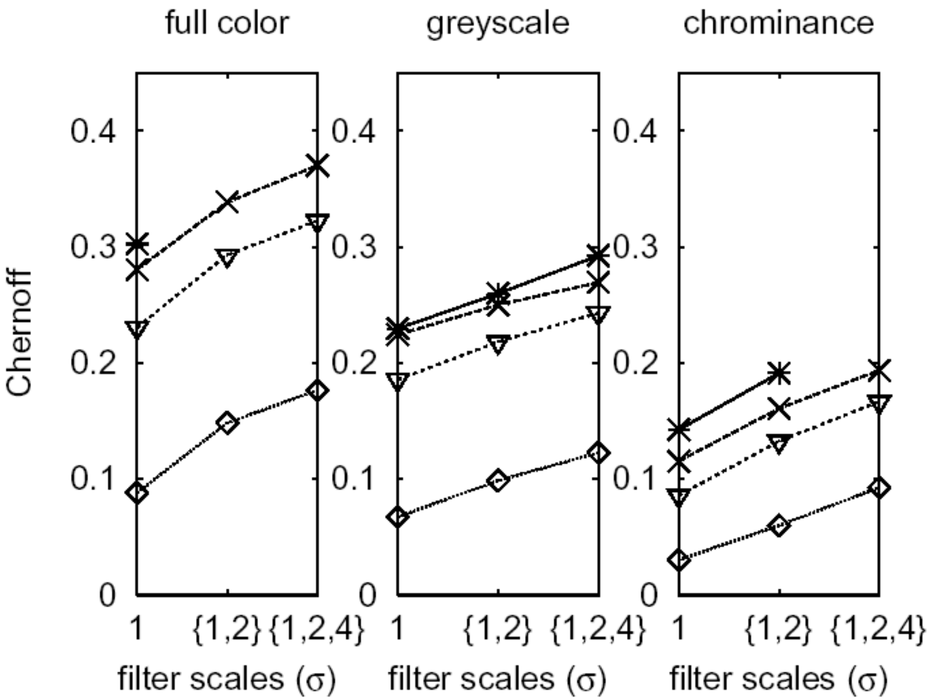
- Gabor filterbank/Hilbert filterbank.

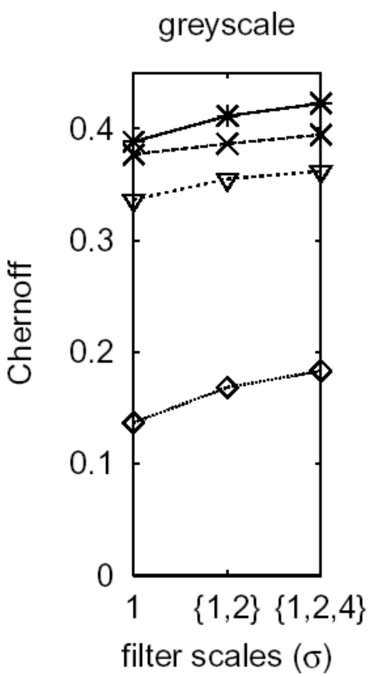


Multiscale

- Sowerby (left), Florida (far right)

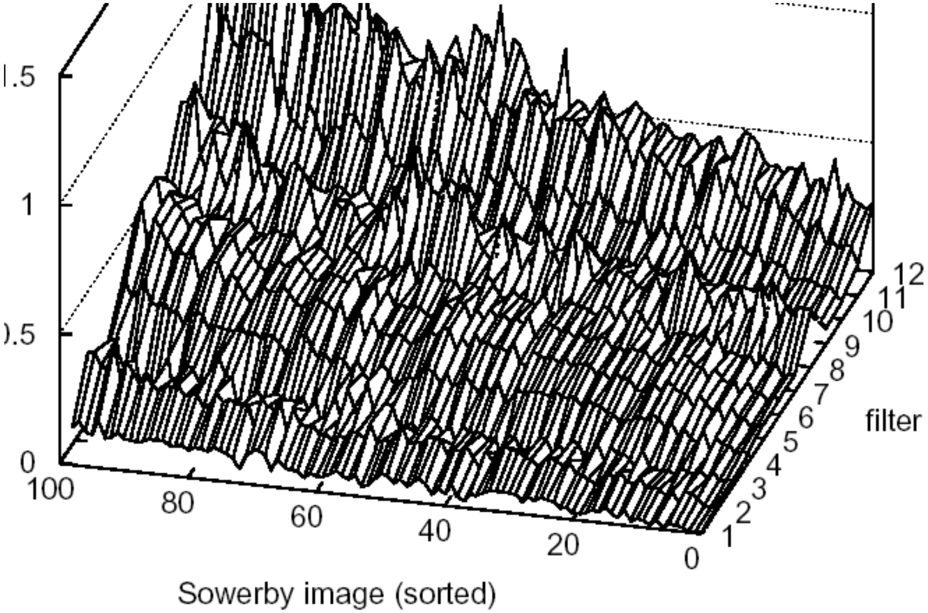
Notation:{1,2} – joints at scales 1,2.

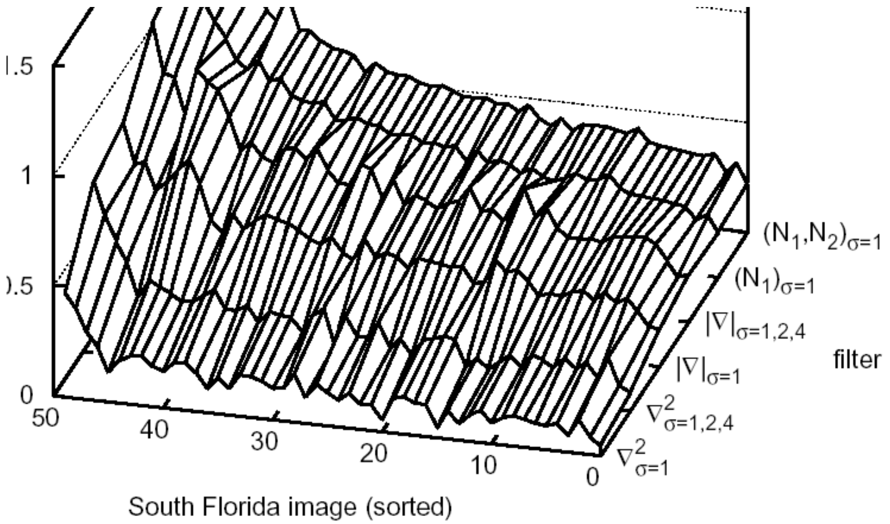




Variations between images:I

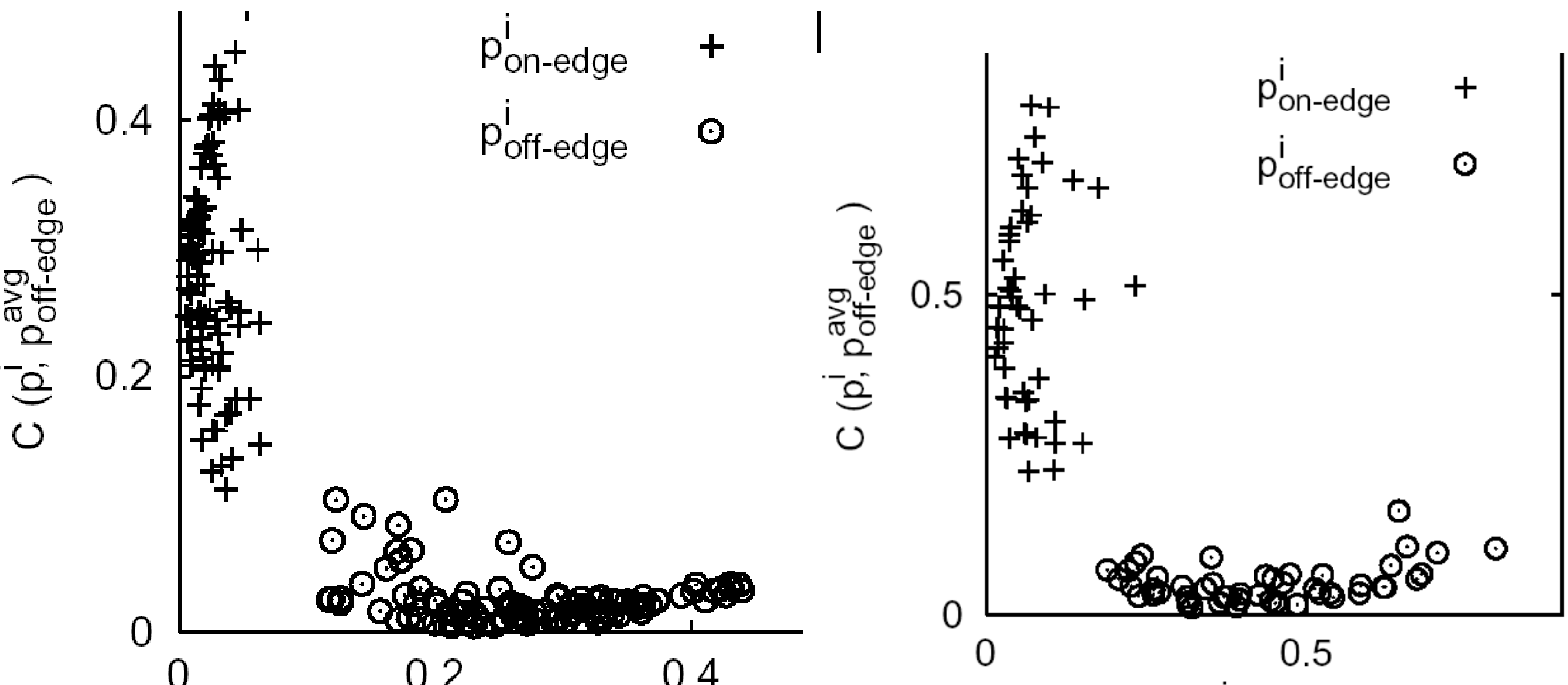
- Relative effectiveness of filter combinations is consistent over dataset.





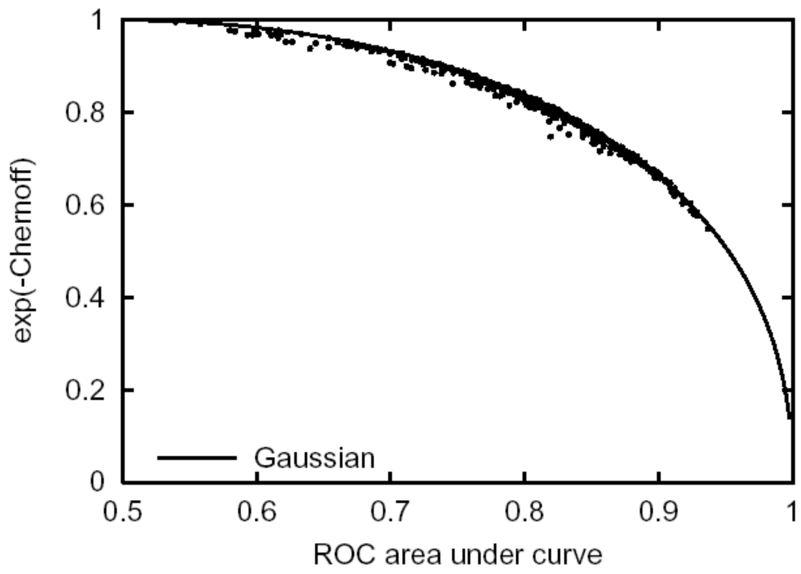
Variations between Images:II

- P(f=y|on) and P(f=y|off) are similar between images. Chernoffs wrt average.



Chernoff and ROC

- Conjectured relationship between Chernoff and ROC (exact for Gaussians). Induced dist. On log-likelihood.



Compare w. Edge Detectors

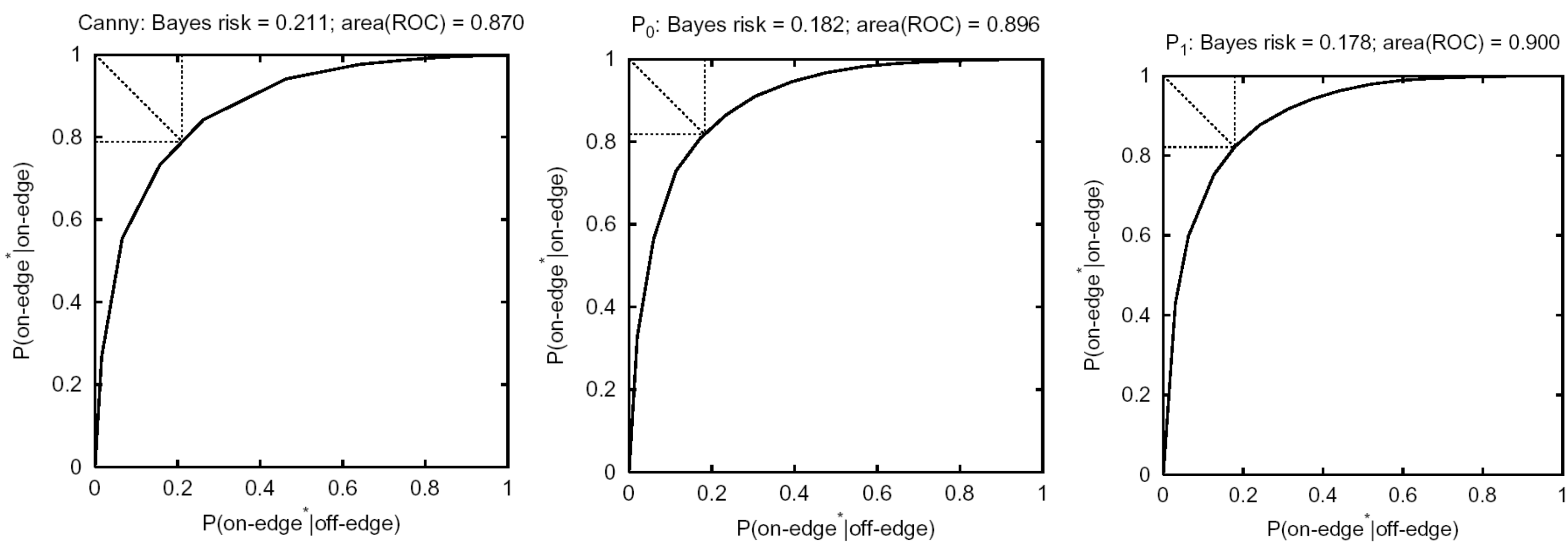
- (I). Florida Dataset.
- Bowyer et al. (2000) evaluated 8 edge detectors. Bayes risk in range 0.035-0.045.

Statistical edge detection gave Bayes risk 0.0350. Canny at 0.0352 (our implement)

Note: little texture/clutter in Florida. Edges at single scale (small scale filters most effective).

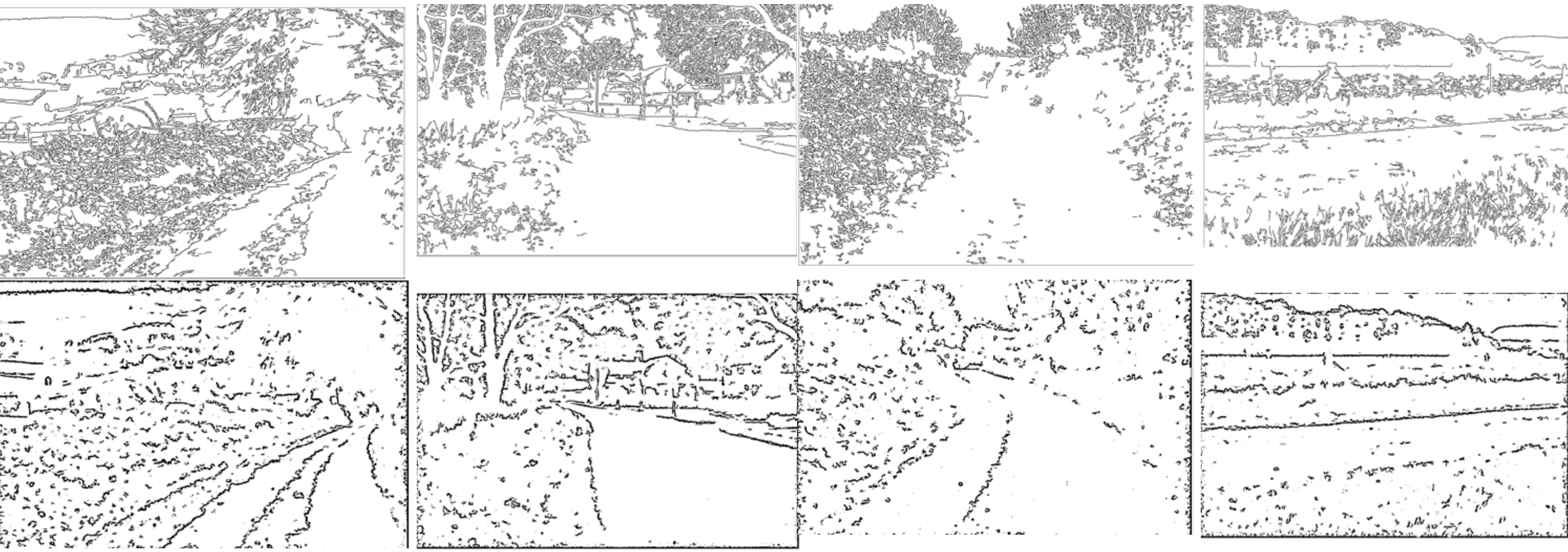
Compare w. Edge Detectors

- (II) Sowerby. More texture/clutter and edges at multiple scales. Statistical edge Detection (right) outperforms Canny (left)



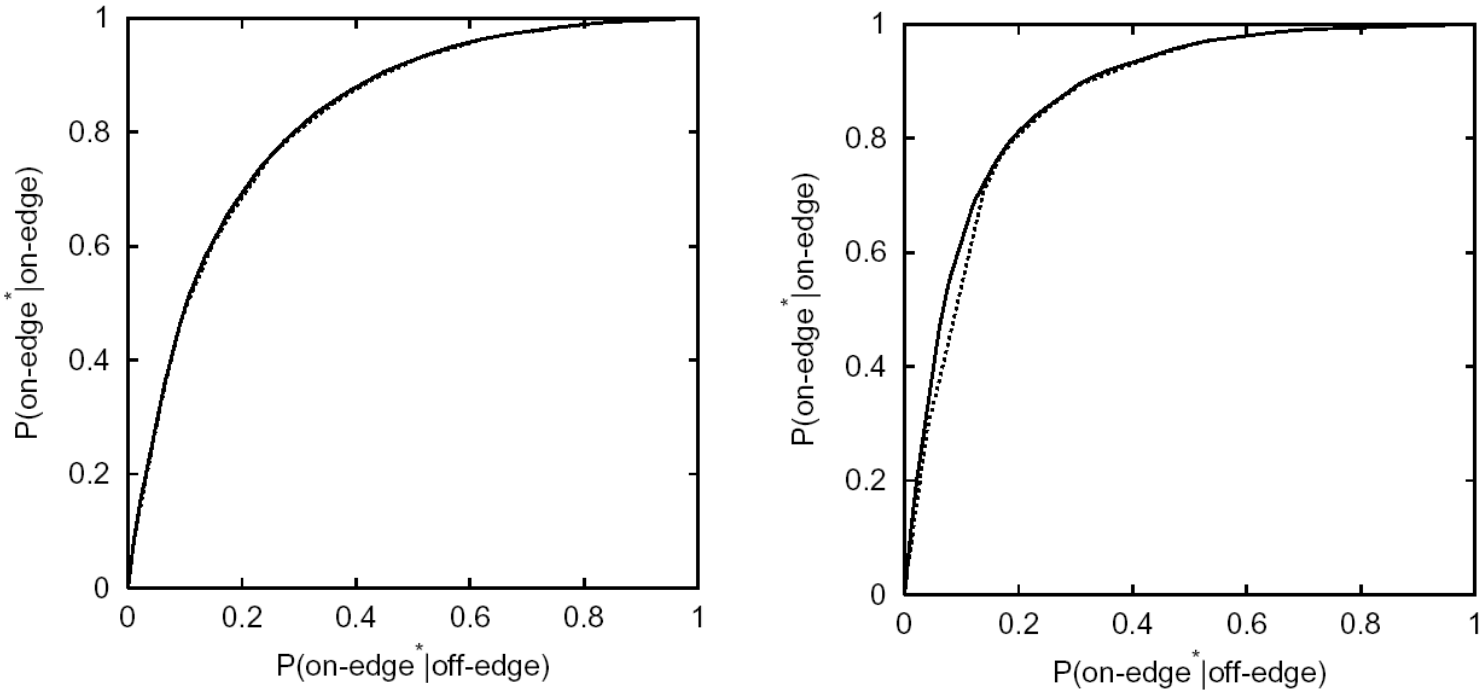
Compare w. Canny

- Canny (top), Statistics (bottom).



Adaption – Sowerby & Florida

- Learn stats on one dataset and adapt to the other. (Scaling assumption).



Extras:

- Localization: Multiple classification: on edge, 1 pixel from edge, 2 pixels, etc.

(Konishi, Yuille, Coughlan 2002).

Region Identification: Vegetation, Sky, Road, Building, etc.

(Konishi and Yuille 1999).

Summary

(I) Statistical regularities of ON and OFF edge. (Extends studies of image stats.)

(II) Implemented a Statistical Edge Detector on 2 datasets – showed it outperformed alternatives quantitatively.

(III) Easy to combine with other stat algs.

(IV) There are many stat. regs. in images.